SECURITY CLASSIFICATION OF THIS PAGE

DTIC ELECTE JUN 0 1 1989 S D

AD-A208 272

1a REPORT SECURITY CLASSIFICATION: (U)	1b
2a SECURITY CLASSIFICATION AUTHORITY: N/A	3 DISTRIBUTION/AVAILABILITY OF REPORT: Distribution unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): University of Maryland, College Park	5 MONITORING ORGANIZATION REPORT NUMBER(S): N/A
6a NAME OF PERFORMING ORGANIZATION: University of Maryland	6b OFFICE SYMBOL (If applicable): N/A
6c ADDRESS (City, State, and ZIP Code): Department of Chemistry & Biochemistry, College Park, MD 20742	7a NAME OF MONITORING ORGANIZATION: Office of Naval Research
	7b ADDRESS (City, State, and ZIP Code): 800 N. Quincy Street, Arlington, VA 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION: Office of Naval Research	8b OFFICE SYMBOL (If applicable): ONR
	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-88-K-0323
8c ADDRESS (City, State, and ZIP Code): 800 North Quincy Street, Arlington, VA 22217-5000	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO 61153N; PROJECT NO 1141MB; TASK NO 441N009--01; WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
Macromolecular Calculations for the XTAL-System of Crystallographic Programs

12 PERSONAL AUTHOR(S)
Stewart, James M.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Annual	FROM 6/1/88 TO 5/31/89	1989, June, 1	12

16 SUPPLEMENTARY NOTATION
Prepared in collaboration with K. B. Ward and D. M. Collins at the Naval Research Laboratory (Code 6030). See attached summary and preprints.

17 COSATI CODES		
FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Macromolecular crystallography, Protein crystallography, Crystallographic computing, Computing for molecular biology. (mom)

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

During the last year effort has been concentrated on the adaptation of the constrained refinement programs. PROTIN & PROLSQ to the XTAL system of crystallographic programs. The checking and documentation of programs for maximum entropy density calculations has been completed, programs MERUN, MEDENS, MEFFIT. Two new programs for modifying and manipulating electron density functions, CONVOL and FOGNU have been coded and documented. The generation of a test deck of idealized macromolecular data has been undertaken.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	(U)
22a NAME OF RESPONSIBLE INDIVIDUAL: Dr. Michael Marron	22b TELEPHONE (Include Area Code): 202-696-4760 / 22c OFFICE SYMBOL: ONR

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

89 5 30 160

DATE: 19 May 1989

ANNUAL REPORT ON CONTRACT ONR N00014-88-K-0323

PRINCIPAL INVESTIGATOR: James M. Stewart

CONTRACTOR: University of Maryland - Code 07419

CONTRACT TITLE: Macromolecular Calculations for the XTAL System of Crystallographic Computer Programs

CONTRACT PERIOD: 1988 June 1 through 1991 May 31

RESEARCH OBJECTIVE: To produce, within the XTAL system of crystallographic computer programs, codes, documentation and tests for macromolecular crystallographic calculations.

This project is being carried out in collaboration with K. B. Ward and D. M. Collins of the Naval Research Laboratory (Code 6030).

During the last year progress has been made on the implementation of refinement programs for macromolecular crystals. In addition, documentation has been generated for programs completed during the year for inclusion in the XTAL version 2.6 manual. The following brief description indicates the nature of the programs currently being checked and documented:

PROTIN: A program to read "constraint information" to describe the elements of a macromolecular crystal structure. This information, for example, can be the idealized coordinates for all the amino acids. This program is a translation of the Hendrickson-Konnert program by the same name. The XTAL version is coded and checked to some extent and preliminary documentation has been prepared.

PROLSQ: A program to do constrained least-squarers refinement of the positional and thermal parameters of a macromolecular structure. This program is a translation into XTAL of the Konnert-Hendrickson program by the same name. The XTAL version is coded and at the present date checkout has begun.

Whereas PROTIN and PROLSQ are presently still in check out and documentation preparation, preprints of the descriptive documentation for the other programs which have been completed this year are attached to this report. These preprints contain descriptions of the programs and their purpose. These are for the programs MERUN, MEDENS, MEFFIT, CONVOL, and FOGNU. In addition, to the work on these links in the system work has been done on analyzing problems

with the preexisting XTAL program MIR which produces macromolecular phase information from multiple isomorphous diffraction data. Work has also been started on a test case using ideal synthetic data for the purpose of verifying the correctness of the programs written for macromolecular structure determination.

The following preprints are parts of chapters in the book "XTAL 2.6 User's Manual" by S. R. Hall and J. M. Stewart published by the Department of Crystallography, University of Western Australia.

CONVOL: Real Space Convolution By Multiplication of Transforms

Douglas M. Collins; Laboratory for the Structure of Matter 6030
Naval Research Laboratory, Washington D.C., 20375

Trina M. Norden & James M. Stewart: Department of Chem/Biochem
University of Maryland, College Park, MD 20742

SYNOPSIS

CONVOL provides a means for the real space convolution or "faltung" of density with another function by multiplication of their respective structure factors in reciprocal space. The functions available in CONVOL are a subset of those given in the International tables for X-Ray Crystallography, Vol. II, Mathematical Tables (1967) Table 2.5.3D page 72. Structure factors from an input BDF are extracted, multplied by the chosen transform, and the output BDF contains the products as structure factors. Fourier synthesis of these new structure factors gives a function which may be operationally described as the pointwise replacement of density by a weighted average; for each point the average is over the entire unit cell, and the relative shapeand placement of the weight function is chosen by the user.

METHOD

Convolution, which the program name implies, is not explicitly carried out by CONVOL. Explicit realization of the convolution is achieved through Fourier synthesis of coefficients prepared in CONVOL by multiplication of (complex) structure factors and corresponding values of the transform selected by the user. The transform values are computed from the formulas given in the International Tables and, with one exception, require only normal algebra and FORTRAN intrinsic functions. The exception is a first order Bessel function evaluated through the series expansions given by Abramowitz and Stegun.

Input consists of a command line for the desired function from Table 2.5.3D of the International Tables. This command line also gives the necessary parameters of the function which are indicated in the Table 2.5.3D column headed F(u). An input binary data file is required which contains the observed |Fo| and phase values to be modified. Output is a modified BDF with the modified structure factor and phase values in Fc ready for Fourier transformation. It is important to note that in order to produce the modified map the program FOURR must be used with "FCAL" as the specified coefficient type.

To see the use of this program consider, for example, function 3.1. The application of this function corresponds to the direct space process of density smoothing by local averaging. In direct space the density would be modified by finding, for each point in the map, the local average over points within the surrounding, specified, parallelapiped. Function 3.1 produces the equivalent result by modifying the amplitude and phase of each reflection and then carrying out the Fourier transform. This process requires only a reflection by reflection multiplication in reciprocal space instead of the repeated point by point averaging required in the direct space method.

WARNING: As in the International Tables, parallel coordinates referred to the crystal lattice are used without regard to the possible nonorthogonality of the base. The coordinates are fractional coordinates.

THE CONVOLUTION FUNCTIONS

In a direct space convolution the effect of the programmed function F1.2 is to shift the original density to a new origin at x1, x2, x3.
Required: x1, x2, x3.

In a direct space convolution the effect of the other programmed functions is to replace the original density with a weighted average centered at that point. The weighting functions are:

F2.2 An ellipsoidal weight exponentially decreasing with distance. The general effect is similar to applying an overall anisotropic thermal parameter to smear the density. For a factor b(i), the weight function is w(i)=exp{-(pi**1/3)*[x(i)/b(i)]**2}/b(i) along axis i; the complete weight function is w(1)*w(2)*w(3). Required: b1, b2, b3.

F3.1 A parallelepiped of constant value a in its interior, 0 elsewhere. the size of the parallelepiped is set by lengths b1, b2, b3 parallel to the unit cell edges. The parallelepiped interior is defined by |x(i)| less than b(i). Required: a, b1, b2, b3.

F3.2 Similar to F3.1, but the volume is a cylinder parallel to a crystal axis specified by the user. The cylinder is bounded by a parallelepiped with |x(i)| less than b(i). The formula takes b1 as along the cylinder axis, and reorientation of the cylinder is accomplished by tying each b(i) to a crystal axis. Required: a, b1, b2, b3, and axis specification.

F3.3 An ellipsoid of constant value a in its interior, 0 elsewhere. Along each axis the ellipsoid is bounded by |x(i)| less than b(i). This figure can be a sphere if the base is orthogonal and b1, b2, b3 are properly chosen. Required: a, b1, b2, b3.

F4 An ellipsoid of the same shape as for F3.3, but its value varies inearly with distance from the origin. At its origin the ellipsoid has value a, at its boundary it has the value 0. At an interior point x1, x2, x3, the value is a*(1-|r|); r**2 is the sum of [x(i)/b(i)]**2. Required: a, b1, b2, b3.

F5 The same as F4 except that the value of an interior point is given by a*(1-r**2). Required: a, b1, b2, b3.CONVOL:s similar to F3.3 but the weighting is quadratically ramped.

REFERENCES

International Tables for X-Ray Crystallography, Vol. II, (1967), sections 2.5.3.4 and 6.3, Edited by J. S. Kasper and K. Lonsdale, The Kynoch Press 1967

Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Edited by M. Abramowitz and I. A. Stegun, Dover Publications, Inc. 1965

FOGNU: Generation of Any Portion of an Electron Density Map From a Map of the Whole Unit Cell

Nick Spadacinni, Crystallography Centre, University of Western Australia Nedlands, 6009, AUSTRALIA

Trina M. Norden & James M. Stewart: Department of Chem/Biochem University of Maryland, College Park, MD 20742

SYNOPSIS

FOGNU produces any volume of space from an electron density map of the modulo one unit cell. The output map will have the same grid and orientation as the input map, only the limits of the map may be changed according to user commands. The map may be output as a new binary file readable by MAKBRK, CONTRS, etc. or printed or both.

METHOD

The method is to read the input map from file C, line by line, generating any additional points required on the line and writing each expanded line to the output as it is formed. Thus maps of very large size may be handied. The limitation is only in terms of the the maximum number of points in the third sum direction. This number cannot exceed the value of the XTAL macro binseqbuf:, an installation specific value usually greater than 500.

MEDENS: Compute the Constrained Exponential Electron Density

Douglas M. Collins: Laboratory for the Structure of Matter 6030
Naval Research Laboratory, Washington D.C. 20375

James M. Stewart: Department of Chemistry/Biochemistry
University of Maryland, College Park, MD 20742

SYNOPSIS

MEDENS takes in a low resolution direct space (electron density) map and computes the constrained exponential electron density distribution. The resulting density function is written to output. The output function is a function of maximum entropy and is intended to be used for the extrapolation, interpolation and smoothing of reflection phases. The resulting phases are recovered by the use of the XTAL program RFOURR. MEDENS may be used most easily by means of the program MERUN which sets up the run stream for all the links necessary in a maximum entropy refinement. This program is modeled on a subroutine "MAXENT" written by E. Prince at the National Bureau of Standards, Gaithersburg, Maryland.

PURPOSE

The MEDENS program processes an electron density map stored on a scratch BDF of the kind generated by XTAL program FOURR. This map will usually be an Fo Fourier transform produced using low resolution phase data. The output file will be the same form as the input file and all the electron density will have been processed to be a function of maximum entropy. This means that new density will conform to an exponential distribution. All negative areas will have been scaled above zero and all large positive regions will have been sharpened. The final sharpness may be controlled by the use of a sharpening parameter under the users control. Values of the sharpening parameter greater than one enhance the sharpness of the map, while those less damp it. The MEDENS process

provides an electron density map, that when processed by RFOURR, will produce phases for more reflections than were used to create the input map. The input map must be for a whole unit cell and done at sufficient resolution to allow the phase extension.

METHOD

The input electron density is scanned to establish the maximum, minimum and average electron density. The process that follows is very sensitive to average electron density, and will not work if the F(0,0,0) term has been left out of the original electron density calculation. The process becomes ill-conditioned and will fail as the mean of the electron density approaches zero from the positive. It is also important that the grid of the input electron density be "fine" enough to provide resolution that will allow the phase extension in the subsequent RFOURR run and that the chosen grid sizes in the three crystallographic directions conform to the FFT grid restrictions for that program.

The output electron density with unit sharpening is an exponential representation which satisfies two constraints:

1) The mean of the electron density remains constant.

2) The mean square of the electron density remains constant.

This is accomplished by calculating the new electron density, NED, from the old electron density, OED, by:

$$NED = EXP(Z*B+(1 - Z)*A)$$

where $$Z = (OED - MIN(OED))/(MAX(OED) - MIN(OED))$$

The scale factors A and B are obtained by a Newton-Raphson iteration.

The initial values of A and B are:

$$A = Ln(0.005*MAX(OED))$$
$$B = Ln(MAX(OED))$$

Using the values of A and B at each iteration four summations are made:

1) The sum of all NED (This is constraint 1)
2) The sum of all NED**2 (This is constraint 2)
3) The sum of all Z*NED
4) The sum of all Z*NED**2

During the initial survey the sums of all OED and OED**2 are saved in

order to calculate scale and tolerance factors.

A renormalization scale factor is obtained by:

$$S = \text{sum(OED)/sum(NED)}$$

and two tolerance (agreement) factors are calculated:

$$T' = \text{ABS(sum(NED) - sum(OED))/sum(OED)}$$
$$T'' = \text{ABS((sum(NED**2)*(S**2) - sum(OED**2))/sum(OED**2)}$$

when the larger of these two factors is less than 0.001, the iteration procedure is stopped. At each iteration A and B are adjusted by:

$$A' = A + \text{Ln}(S)$$
$$B' = B + \text{Ln}(S) + Q'$$

where $Q' = \text{(sum(OED**2) - (S**2)*(sum(NED**2)))/D}$
where $D = E*F$
where $E = 2.0*S*\text{sum(OED)}*F$
where $F = \text{sum(Z*NED**2)/sum(NED)} - G$
where $G = \text{((sum(NED**2)*(sum(Z*NED))/(sum(NED))**2}$

The value of Q' is always constrained to lie between plus and minus LL, where LL is 1.0 by default, but may be set by the user.

Once convergence has been achieved, the new electron density is computed from the final values of A and B. The new electron density is written to output on file D in the same format as it came from file C created by FOURR.

The final electron density is calculated by:

$$Z = \text{(OED - MIN(OED))/(MAX(OED) - MIN(OED))}$$
$$NED = S*\text{EXP(SHARPF*(Z*B + (1.0 - Z)*A)}$$

SHARPF is a parameter which controls the sharpness of the output map. The default value is 1.0, however, in many applications, 0.5, corresponding to the square root, may be found to be more satisfactory. If the value of SHARPF is set in the MEDENS line, the final density is rescaled to restore the correct mean value, but the original mean square of the density is not recovered.

REFERENCE

Phase Extension by Combined Entropy Maximization and Solvent Flattening, E. Prince, L. Sjolin, and R. Alenljung Acta Cryst. (1988). A44, 216-222

MEFFIT: Modify Constrained Electron Density

Douglas M. Collins; Laboratory for the Structure of Matter 6030
Naval Research Laboratory, Washington D.C., 20375

James M. Stewart: Department of Chemistry/Biochemistry
University of Maryland, College Park, MD 20742

SYNOPSIS

MEFFIT combines a positive definite map of constrained exponential electron density formed by MEDENS and a difference map prepared by FOURR using phases calculated by RFOURR from the MEDENS output map. The effect of this process is to produce a new map such that the maximum entropy phases and the observed structure moduli will be produced upon Fourier transformation of the new map. The program may be run by use of the XTAL program MERUN which sets up an input stream to drive all the programs needed in the refinement process.

METHOD

The focus of program MEFFIT is adjustment of a positive definite density map toward agreement with observed structure factor magnitudes. The designed application is sequential improvement of an imperfectly phased set of structure factors through manipulation of its corresponding noisy or low resolution density function by MEDENS, MEFFIT, and necessary Fourier routines. On a grid suitably fine for the desired final resolution, an initial electron density function is converted by MEDENS into a maximum-entropy density, which is a positive definite exponential function. This map is used as input to RFOURR to calculate structure factors (including phases), which, it is expected, are subsequently used in FOURR to form a companion difference electron density map.

MEFFIT generates a new positive-definite exponential density in a maximum-entropy adjustment of the prior exponential density. The new density is computed by pointwise multiplication of the former by exp{constant*difference density}, then scaled to restore the original mean value. As a stand alone process, this combination of a positive definite

density and the difference density which it implies, constitutes one maximum-entropy step in the adjustment
of a positive definite density to more closely match the experimentally observed moduli.

A useful number of MEFFIT steps is likely to be in the range 4 - 10. The associated file handling of prior, difference, and updated maps, in addition to the usual BDF handling, is simple but a fruitful source of confusion. For the usual iterative application of MEFFIT, MERUN can be used to construct the XTAL input stream and run the programs. Although the criterion for completion might be a target R value which is suitably small, if the R value is changing slowly after several iterations, it is unlikely that further substantive change is occurring in either the structure factor phasing or the exponential density itself. In any case, the process is exceptionally stable, and not critically dependant on the termination point.

After the desired number of iterations have been run, the last part of which is RFOURR to compute structure factors, the final Fobs Fourier map may be calculated. This gives the electron density in its standard formulation, but with structure factor phasing corresponding to the final exponential density. This whole process has the structure: FOURR; MEDENS; RFOURR; (FOURR; MEFFIT; RFOURR); FOURR; in which the parentheses show the inner loop of programs.

REFERENCE

Electron Density: An Exponential Model, D.M. Collins and M.C. Mahar
Acta Cryst. (1983). A39, 252-256.

MERUN: Prepare input stream for MEDENS & MEFFIT programs

Douglas M. Collins: Laboratory for the structure of matter
Code 6030, Naval Research Laboratory, Washington D.C. 20375

James M. Stewart: Department of Chemistry/Biochemistry
University of Maryland, College Park, MD 20742

SYNOPSIS

MERUN takes in specifications for a MEDENS run to produce an initial direct space (electron density) map which is to be refined by MEFFIT. In the cycles of refinement the initial map is adjusted toward compliance with fixed structure moduli and the corresponding phase information is kept as

calculated phases. During the cycles of refinement difference maps will be generated by the FOURR program and new calculated phases and moduli will be produced by RFOURR.

PURPOSE

The program MERUN sets up the XTAL input control lines required to run MEDENS followed by the requested number of cycles of RFOURR; FOURR; MEFFIT; followed by a final RFOURR. The main purpose of the program is to simplify the multiple use of several other XTAL programs and to automate, to a large extent, the necessary I/O file switching.

MERUN facilitates the process of extending and refining phases by the process of taking an initial electron density map and computing the constrained exponential electron density distribution for it by means of the program MEDENS. The output map may then be reverse Fourier transformed by the program RFOURR to produce structure factor phases and moduli corresponding to the maximum entropy map from MEDENS. From these moduli and phases and the observed structure factors it is possible to calculate a difference Fourier map using the program FOURR. This difference map and the original maximum entropy map are combined by use of the program MEFFIT to produce a modified maximum entropy map. This new map, when reverse Fourier transformed, will show moduli in better agreement with the observed moduli. If the initial phase set was of limited resolution, the resolution of each successive map will be higher at a rate dependent upon the acceptance criteria set for MEFFIT and RFOURR. In general several cycles will be required to achieve good agreement between the calculated and observed moduli and to expand fully their phasing and the resolution of the map.

METHOD

The method, as far as MERUN is concerned, consists of forming input line images and placing them in the "punch" file. These images are either generated by default or may be passed through as model images from the MERUN input. The code generates the necessary FILES and END lines.

The input stream generated by MERUN may be automatically added to the XTAL input stream or may simply be printed for inspection. See also the MEDENS and MEFFIT write-ups for the basis for these calculations.